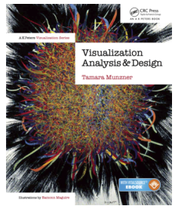
Lectures 3&4:
Facet & Reduce

Tamara Munzner
Department of Computer Science
University of British Columbia



DSCI 532, Data Visualization 2
Week 2, Jan 9 / Jan 11 2018

www.cs.ubc.ca/~tmm/courses/mds-viz2-17

@tamaramunzner

Facet Into Multiple Views

Facet

- Juxtapose

- Partition

- Superimpose

Juxtapose and coordinate views

- Share Encoding: Same/Different
  - *Linked Highlighting*

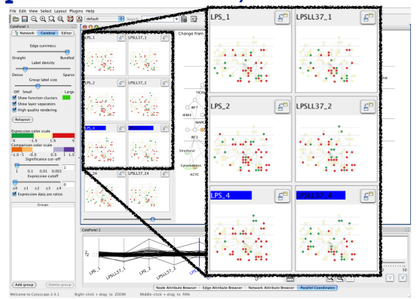
- Share Data: All/Subset/None

- Share Navigation

Idiom: **Small multiples**

System: **Cerebral**

- encoding: same
- data: none shared
  - –different attributes for node colors
  - –(same network layout)
- navigation: shared



[Cerebral:Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

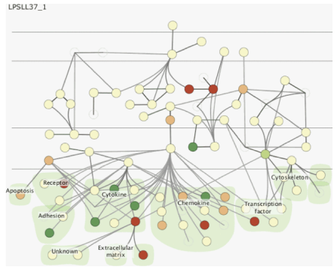
Coordinate views: Design choice interaction

		Data: All	Data: Subset	Data: None
Encoding: Same		Redundant	Overview/ Detail	Small Multiples
Encoding: Different		Multiform	Multiform, Overview/ Detail	No Linkage

- why juxtapose views?
  - –benefits: eyes vs memory
    - lower cognitive load to move eyes between 2 views than remembering previous state with single changing view
  - –costs: display area, 2 views side by side each have only half the area of one view

Why not animation?

- disparate frames and regions: comparison difficult
  - –vs contiguous frames
  - –vs small region
  - –vs coherent motion of group
- safe special case
  - –animated transitions



Eyes beat memory

- principle: external cognition vs. internal memory
  - –easy to compare by moving eyes between side-by-side views
  - –harder to compare visible item to memory of what you saw
- implications for animation
  - –great for choreographed storytelling
  - –great for transitions between two states
  - –poor for many states with changes everywhere
    - consider small multiples instead

literal — abstract

animation — small multiples

show time with time — show time with space

Change blindness

- if attention is directed elsewhere, even drastic changes not noticeable
  - –door experiment
- change blindness demos
  - –mask in between images

Idiom: **Linked highlighting**

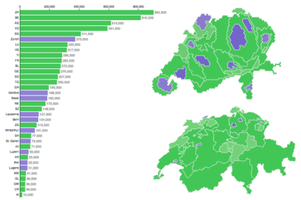
System: **EDV**

- see how regions contiguous in one view are distributed within another
  - –powerful and pervasive interaction idiom
- encoding: different
  - –***multiform***
- data: all shared
- aka: brushing and linking

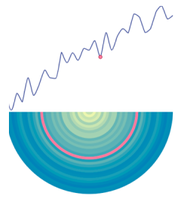
[Visual Exploration of Large Structured Datasets. Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237–246. IOS Press, 1995.]

Linked views

- unidirectional vs bidirectional linking



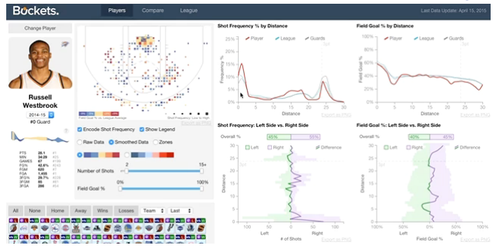
http://www.ralphstraumann.ch/projects/swiss-population-cartogram/



http://peterbeshai.com/linked-highlighting-react-d3-reflux/

Linked views: Multidirectional linking

System: **Buckets**



http://buckets.peterbeshai.com/

https://medium.com/@pbesh/linked-highlighting-with-react-d3-js-and-reflux-16e9c0b2210b

Idiom: **Overview-detail views**

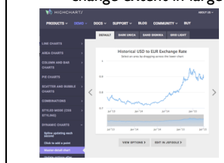
System: **Google Maps**

- encoding: same
- data: subset shared
- navigation: shared
  - –bidirectional linking
- differences
  - –viewpoint
  - –(size)
- special case: ***birds-eye map***

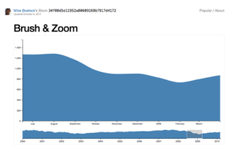
[A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.]

Idiom: **Overview-detail navigation**

- encoding: same
- data: subset shared
- navigation: shared
  - –unidirectional linking
  - –select in small overview
  - –change extent in large detail view



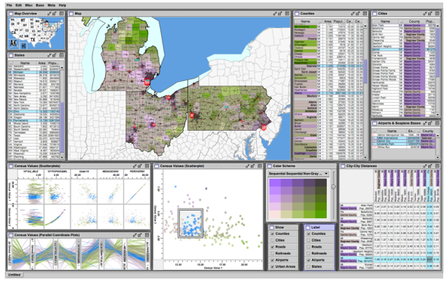
https://www.highcharts.com/demo/dynamic-master-detail



https://bl.ocks.org/mbostock/34f08d5e11952a80609169b7917d4172

System: **Improvise**

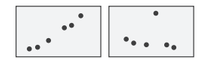
- investigate power of multiple views
  - –pushing limits on view count, interaction complexity
  - –how many is ok?
    - open research question
  - –reorderable lists
    - easy lookup
    - useful when linked to other encodings



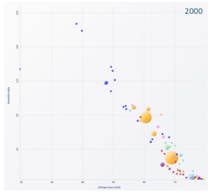
[Building Highly-Coordinated Visualizations In Improvise. Weaver. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 159–166, 2004.]

Partition into views

- how to divide data between views → Partition into Side-by-Side Views
  - –split into regions by attributes
  - –encodes association between items using spatial proximity
  - –order of splits has major implications for what patterns are visible

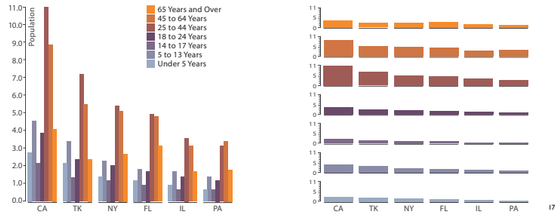


- no strict dividing line
  - –***view:*** *big/detailed*
    - contiguous region in which visually encoded data is shown on the display
  - –***glyph:*** *small/iconic*
    - object with internal structure that arises from multiple marks



Partitioning: List alignment

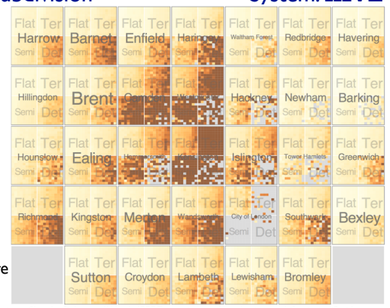
- single bar chart with grouped bars
  - –split by state into regions
    - complex glyph within each region showing all ages
  - –compare: easy within state, hard across ages
- small-multiple bar charts
  - –split by age into regions
    - one chart per region
  - –compare: easy within age, harder across states



Partitioning: Recursive subdivision

System: **HIVE**

- split by neighborhood
- then by type
- then time
  - –years as rows
  - –months as columns
- color by price
- neighborhood patterns
  - –where it's expensive
  - –where you pay much more for detached type



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Partitioning: Recursive subdivision

System: **HIVE**

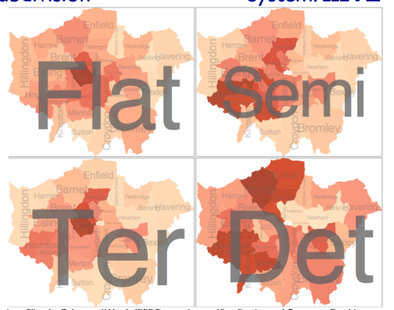
- switch order of splits
  - –type then neighborhood
- switch color
  - –by price variation
- type patterns
  - –within specific type, which neighborhoods inconsistent

[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Partitioning: Recursive subdivision

System: **HIVE**

- different encoding for second-level regions
  - –choropleth maps

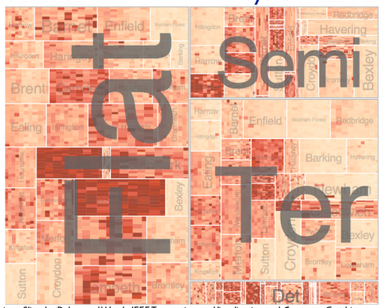


[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Partitioning: Recursive subdivision

System: **HIVE**

- size regions by sale counts
  - –not uniformly
- result: treemap



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Superimpose layers

- ***layer***: set of objects spread out over region
  - –each set is visually distinguishable group
  - –extent: whole view
- design choices
  - –how many layers, how to distinguish?
    - encode with different, nonoverlapping channels
    - two layers achieveable, three with careful design
  - –small static set, or dynamic from many possible?

Superimpose Layers

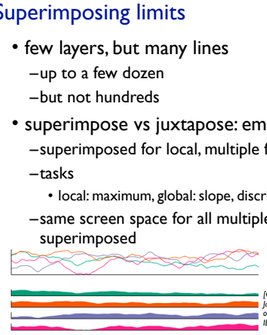
Static visual layering

- foreground layer: roads
  - –hue, size distinguishing main from minor
  - –high luminance contrast from background
- background layer: regions
  - –desaturated colors for water, parks, land areas
- user can selectively focus attention
- "get it right in black and white"
  - –check luminance contrast with greyscale view

[Get it right in black and white. Stone. 2010. http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white]

Superimposing limits

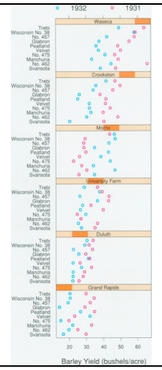
- few layers, but many lines
  - –up to a few dozen
  - –but not hundreds
- superimpose vs juxtapose: empirical study
  - –superimposed for local, multiple for global
  - –tasks
    - local: maximum, global: slope, discrimination
  - –same screen space for all multiples vs single superimposed



[Graphical Perception of Multiple Time Series. Javed, McDonnel, and Elmqvist. IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE InfoVis 2010) 16:6 (2010), 927–934.]

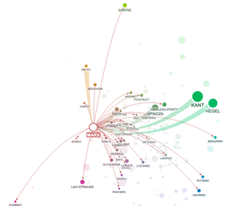
Idiom: **Trellis plots**

- superimpose within same frame
  - –color code by year
- partitioning
  - –split by site, rows are wheat varieties
- main-effects ordering
  - –derive value of median for group, use to order
  - –order rows within view by variety median
  - –order views themselves by site median

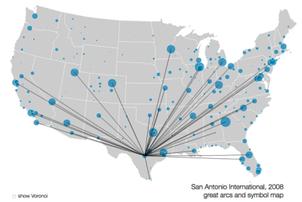


Dynamic visual layering

- interactive based on selection
- one-hop neighbour highlighting demos: click vs hover (lightweight)



http://mariandoerk.de/edgemaps/demo/



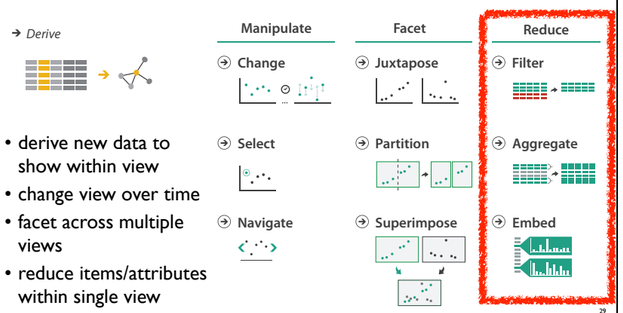
http://mbostock.github.io/d3/talk/20111116/airports.html

Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 12: Facet Into Multiple Views*
- *A Review of Overview+Detail, Zooming, and Focus+Context Interfaces.* Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.
- *A Guide to Visual Multi-Level Interface Design From Synthesis of Empirical Study Evidence.* Lam and Munzner. Synthesis Lectures on Visualization Series, Morgan Claypool, 2010.
- *Zooming versus multiple window interfaces: Cognitive costs of visual comparisons.* Plumlee and Ware. ACM Trans. on Computer-Human Interaction (ToCHI) 13:2 (2006), 179–209.
- *Exploring the Design Space of Composite Visualization.* Javed and Elmqvist. Proc. Pacific Visualization Symp. (PacificVis), pp. 1–9, 2012.
- *Visual Comparison for Information Visualization.* Gleicher, Albers, Walker, Jusufi, Hansen, and Roberts. Information Visualization 10:4 (2011), 289–309.
- *Guidelines for Using Multiple Views in Information Visualizations.* Baldonado, Woodruff, and Kuchinsky. In Proc. ACM Advanced Visual Interfaces (AVI), pp. 110–119, 2000.
- *Cross-Filtered Views for Multidimensional Visual Analysis.* Weaver. IEEE Trans. Visualization and Computer Graphics 16:2 (Proc. InfoVis 2010), 192–204, 2010.
- *Linked Data Views.* Wills. In Handbook of Data Visualization, Computational Statistics, edited by Unwin, Chen, and Härdle, pp. 216–241. Springer-Verlag, 2008.
- *Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications.* Borgo, Kehrer, Chung, Maguire, Laramee, Hauser, Ward, and Chen. In Eurographics State of the Art Reports, pp. 39–63, 2013.

Reduce

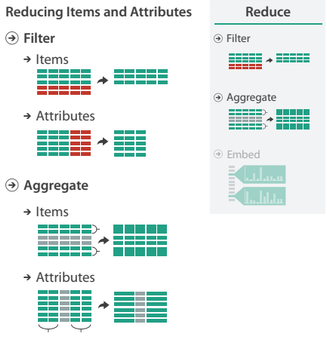
How to handle complexity: 1 previous strategy + 3 more



- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

Reduce items and attributes

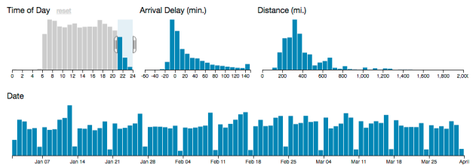
- reduce/increase: inverses
- filter
  - –pro: straightforward and intuitive
    - to understand and compute
  - –con: out of sight, out of mind
- aggregation
  - –pro: inform about whole set
  - –con: difficult to avoid losing signal
- not mutually exclusive
  - –combine filter, aggregate
  - –combine reduce, change, facet



Idiom: **cross filtering**

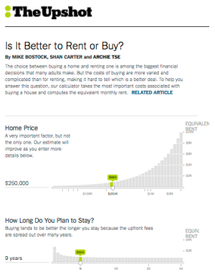
System: **Crossfilter**

- item filtering
- coordinated views/controls combined
  - all scented histogram bisliders update when any ranges change



[http://square.github.io/crossfilter/]

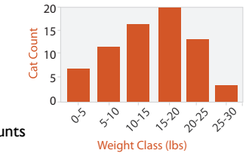
Idiom: **cross filtering**



[https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0]

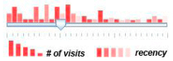
Idiom: **histogram**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly

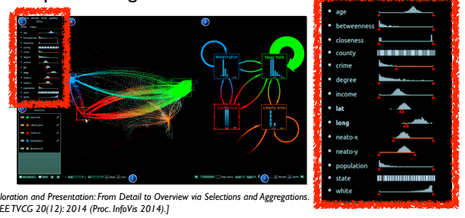


Idiom: **scented widgets**

- augmented widgets show information scent
  - cues to show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider

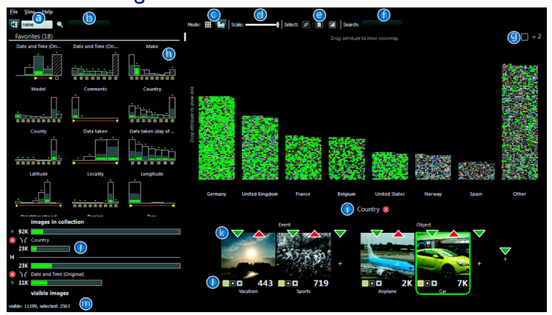


[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]



[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]

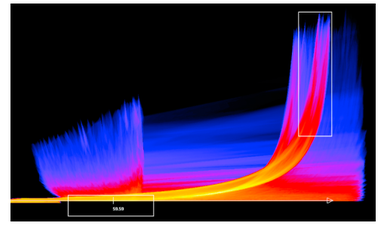
Scented histogram bisliders: detailed



[ICLIC: Interactive categorization of large image collections. van der Corput and van Wijk. Proc. PacificVis 2016.]

Idiom: **Continuous scatterplot**

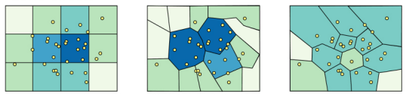
- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap



[Continuous Scatterplots. Bachthaler and Weiskopf. IEEE TVCG (Proc.Vis 08) 14:6 (2008), 1428–1435. 2008.]

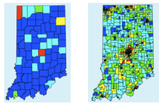
Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - gerrymandering (manipulating voting district boundaries) is only one example!
  - zone effects



[http://www.e-education.psu/edu/geog486/l4_p7.html, Fig 4.cg.6]

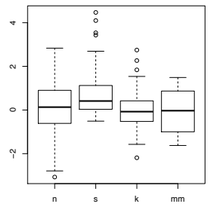
  - scale effects



https://blog.cartographica.com/blog/2011/5/19/the-modifiable-areal-unit-problem-in-gis.html

Idiom: **boxplot**

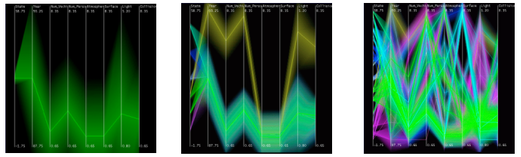
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown



[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Idiom: **Hierarchical parallel coordinates**

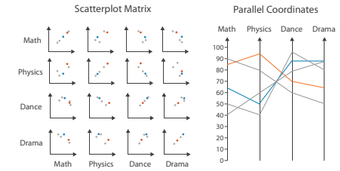
- dynamic item aggregation



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

Idioms: **scatterplot matrix, parallel coordinates**

- scatterplot matrix (SPLOM)
  - rectilinear axes, point mark
  - facet: all possible pairs of axes
  - scalability
    - one dozen attribs
    - dozens to hundreds of items
- parallel coordinates
  - parallel axes, jagged line representing item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
    - hundreds of items



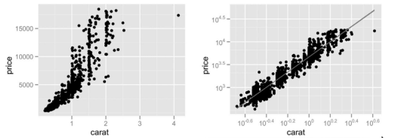
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

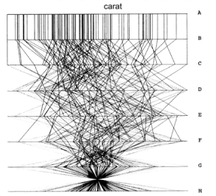
after [Visualization Course Figures. McGuffin, 2014. http://www.michaelmcguffin.com/courses/vis/]

Task: Correlation

- scatterplot matrix
  - positive correlation
    - diagonal low-to-high
  - negative correlation
    - diagonal high-to-low
  - uncorrelated: spread out
- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings



[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

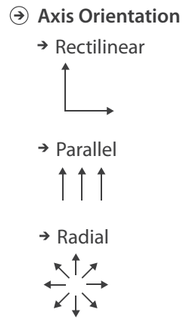


[Hyperdimensional Data Analysis Using Parallel Coordinates. Wegman. Journ. American Statistical Association 85:411 (1990), 664–675.]

Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8$, and -1.

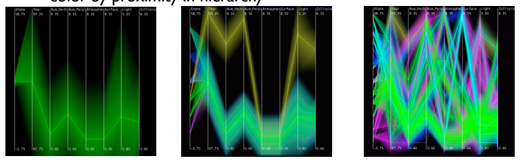
Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best
  - 3 problematic
  - 4+ impossible
- parallel: unfamiliarity, training time



Idiom: **Hierarchical parallel coordinates**

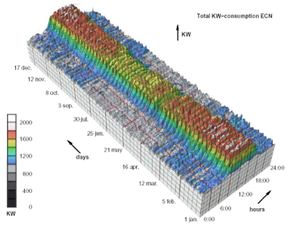
- dynamic item aggregation
- derived data: ***hierarchical clustering***
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

Hierarchical clustering example: time-series data

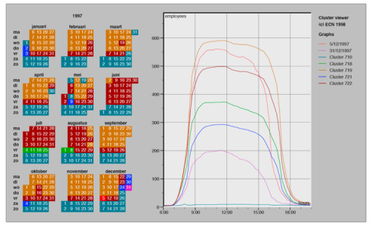
- unjustified 3D with extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Hierarchical clustering example: cluster-calendar

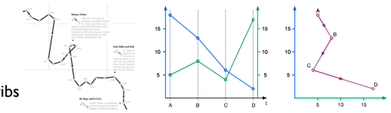
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

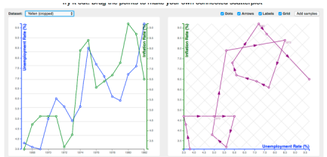


[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Idiom: **connected scatterplots**

- scatterplot with line connection marks
  - popular in journalism
  - horiz + vert axes: value attribs
  - line connection marks: temporal order
  - alternative to dual-axis charts
    - horiz: time
    - vert: two value attribs
- empirical study
  - engaging, but correlation unclear



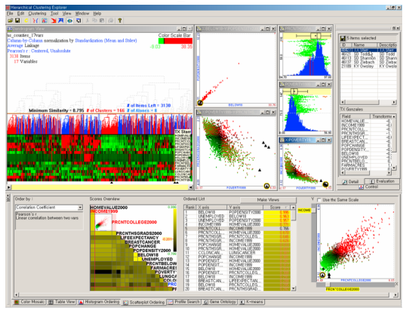


[The Connected Scatterplot for Presenting Paired Time Series. Haroz, Kasara and Franconeri. IEEE TVCG 22(9):2174-86, 2016.]

http://steveharoz.com/research/connected_scatterplot/

System: **Hierarchical Clustering Explorer**

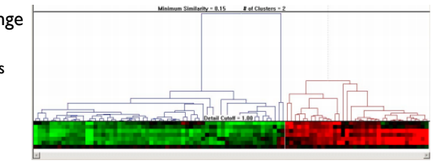
- many linked views
- cluster heatmap
  - dynamic aggregation: hierarchical clustering
  - explicitly visible



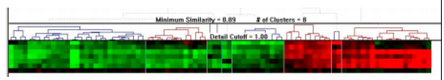
[http://www.cs.umd.edu/hcil/hce/]

System: **Hierarchical Clustering Explorer**

- drag line to change level of detail
  - coarse: 2 clusters



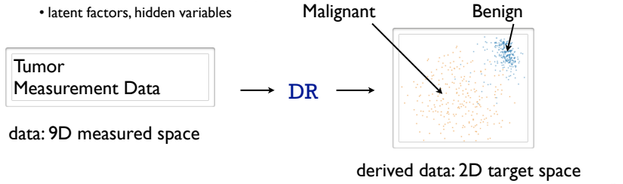
  - fine: 8 clusters



[http://www.cs.umd.edu/hcil/hce/]

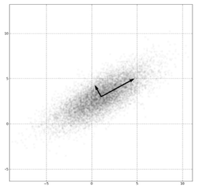
Dimensionality reduction

- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space
    - capture most of variance with minimal error
  - use when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables



Linear dimensionality reduction

- principal components analysis (PCA)
  - finding axes: first with most variance, second with next most, …
  - describe location of each point as linear combination of weights for each axis
    - mapping synthesized dims to original dims



[http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png]

Dimensionality vs attribute reduction

- vocab use in field not consistent
  - dimension/attribute
- attribute reduction: reduce set with filtering
  - includes orthographic projection
- dimensionality reduction: create smaller set of new dims/attribs
  - typically implies dimensional aggregation, not just filtering
  - vocab: projection/mapping

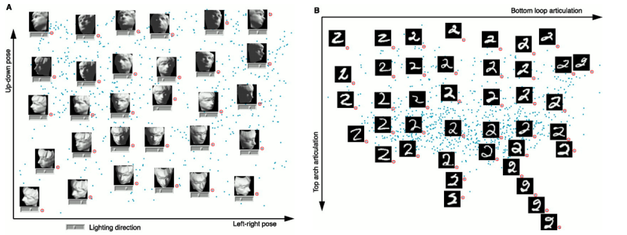
Dimensionality reduction & visualization

- why do people do DR?
  - improve performance of downstream algorithm
    - avoid curse of dimensionality
  - data analysis
    - if look at the output: visual data analysis
- abstract tasks when visualizing DR data
  - dimension-oriented tasks
    - naming synthesized dims, mapping synthesized dims to original dims
  - cluster-oriented tasks
    - verifying clusters, naming clusters, matching clusters and classes

[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Dimension-oriented tasks

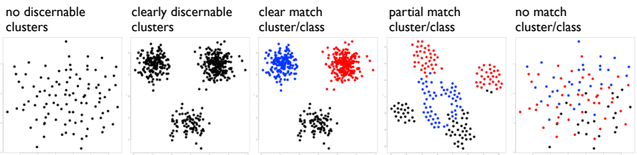
- naming synthesized dims: inspect data represented by lowD points



[A global geometric framework for nonlinear dimensionality reduction. Tenenbaum, de Silva, and Langford. Science, 290(5500):2319–2323, 2000.]

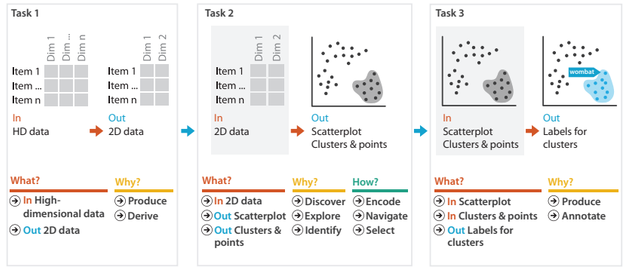
Cluster-oriented tasks

- verifying, naming, matching to classes



[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Idiom: **Dimensionality reduction for documents**



Nonlinear dimensionality reduction

- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attribs
  - new dimensions often cannot be easily related to originals
    - mapping synthesized dims to original dims task is difficult
- many techniques proposed
  - many literatures: visualization, machine learning, optimization, psychology, ...
  - techniques: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE,…
    - t-SNE: excellent for clusters
      - but some trickiness remains: http://distill.pub/2016/misread-tsne/
    - MDS: confusingly, entire family of techniques, both linear and nonlinear
      - minimize stress or strain metrics
      - early formulations equivalent to PCA

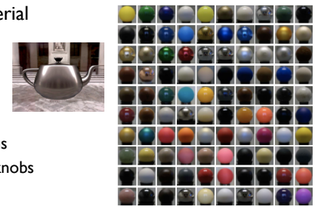
VDA with DR example: nonlinear vs linear

- DR for computer graphics reflectance model
  - goal: simulate how light bounces off materials to make realistic pictures
    - computer graphics: BRDF (reflectance)
  - idea: measure what light does with real materials

[Fig 2. Matusik, Pfister, Brand, and McMillan. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Capturing & using material reflectance

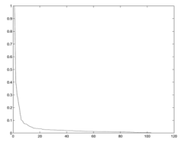
- reflectance measurement: interaction of light with real materials (spheres)
- result: 104 high-res images of material
  - each image 4M pixels
- goal: image synthesis
  - simulate completely new materials
- need for more concise model
  - 104 materials * 4M pixels = 400M dims
  - want concise model with meaningful knobs
    - how shiny/greasy/metallic
    - DR to the rescue!



[Figs 5/6. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Linear DR

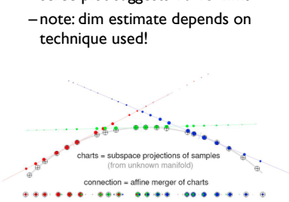
- first try: PCA (linear)
- result: error falls off sharply after ~45 dimensions
  - scree plots: error vs number of dimensions in lowD projection
- problem: physically impossible intermediate points when simulating new materials
  - specular highlights cannot have holes!

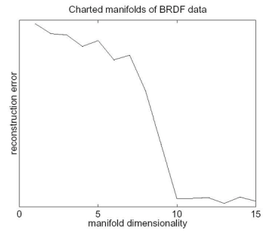


[Figs 6/7. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Nonlinear DR

- second try: charting (nonlinear DR technique)
  - scree plot suggests 10-15 dims
  - note: dim estimate depends on technique used!

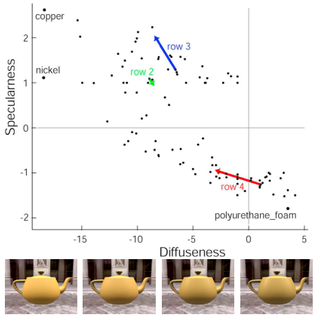




[Fig 10/11. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

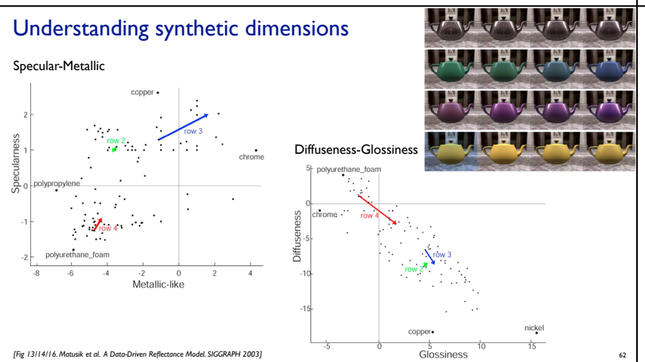
Finding semantics for synthetic dimensions

- look for meaning in scatterplots
  - synthetic dims created by algorithm but named by human analysts
  - points represent real-world images (spheres)
  - people inspect images corresponding to points to decide if axis could have meaningful name
- cross-check meaning
  - arrows show simulated images (teapots) made from model
  - check if those match dimension semantics



[Fig 12/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Understanding synthetic dimensions



[Fig 13/14/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 13: Reduce Items and Attributes*
- *Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines.* Elmqvist and Fekete. IEEE Transactions on Visualization and Computer Graphics 16:3 (2010), 439–454.
- *A Review of Overview+Detail, Zooming, and Focus+Context Interfaces.* Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.
- *A Guide to Visual Multi-Level Interface Design From Synthesis of Empirical Study Evidence.* Lam and Munzner. Synthesis Lectures on Visualization Series, Morgan Claypool, 2010.